

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>5</sup> :</b> <b>C12N 15/70, 1/21, C07H 21/04</b> <b>C07K 13/00, C12P 21/02</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 92/13955</b> <b>(43) International Publication Date:</b> 20 August 1992 (20.08.92)
<b>(21) International Application Number:</b> PCT/US92/00944 <b>(22) International Filing Date:</b> 6 February 1992 (06.02.92)  <b>(30) Priority data:</b> 652,531 6 February 1991 (06.02.91) US 745,382 14 August 1991 (14.08.91) US  <b>(71) Applicant:</b> GENETICS INSTITUTE, INC. [US/US]; 87 CambridgePark Drive, Cambridge, MA 02140 (US).  <b>(72) Inventors:</b> MCCOY, John ; 63 Pine Ridge Road, Reading, MA 01876 (US). LAVALLIE, Edward, R. ; 90 Green Meadow Drive, Tewksbury, MA 01876 (US).		<b>(74) Agent:</b> CSERR, Luann; Legal Affairs, Genetics Institute, Inc., 87 CambridgePark Drive, Cambridge, MA 02140 (US).  <b>(81) Designated States:</b> AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), MC (European patent), NL (European patent), SE (European patent).  <b>Published</b> <i>With international search report.</i>
<b>(54) Title:</b> PEPTIDE AND PROTEIN FUSIONS TO THIOREDOXIN AND THIOREDOXIN-LIKE MOLECULES  <b>(57) Abstract</b>  This invention provides a fusion molecule comprising a DNA sequence encoding a thioredoxin-like protein fused to the DNA sequence encoding a selected heterologous peptide or protein. The peptide or protein may be fused to the amino terminus of the thioredoxin-like molecule, the carboxyl terminus of the thioredoxin-like molecule, or within the thioredoxin-like molecule, for example at the active-site loop of said molecule. Expression of this fusion molecule under the control of a regulatory sequence capable of directing its expression in a desired host cell, produces high levels of stable and soluble fusion protein. The fusion protein, located in the bacterial cytoplasm, may be selectively released from the cell by osmotic shock or freeze/thaw procedures. It may be optionally cleaved to liberate the soluble, correctly folded heterologous protein from the thioredoxin-like portion.		

BEST AVAILABLE COPY

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	ML	Mali
AU	Australia	FR	France	MN	Mongolia
BB	Barbados	GA	Gabon	MR	Mauritania
BE	Belgium	GB	United Kingdom	MW	Malawi
BF	Burkina Faso	GN	Guinea	NI	Netherlands
BG	Bulgaria	GR	Greece	NO	Norway
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IE	Ireland	RO	Romania
CA	Canada	IT	Italy	RU	Russian Federation
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark	MG	Madagascar		
ES	Spain				

PEPTIDE AND PROTEIN FUSIONS TO THIOREDOXIN  
AND THIOREDOXIN-LIKE MOLECULES

5       The invention relates generally to the production of fusion  
proteins in prokaryotic and eukaryotic cells. More specifically,  
the invention relates to the expression in host cells of  
recombinant fusion sequences comprising thioredoxin or  
thioredoxin-like sequences fused to sequences for selected  
10 heterologous peptides or proteins, and the use of such fusion  
molecules to increase the production, activity, stability or  
solubility of recombinant proteins and peptides.

Background of the Invention

15       Many peptides and proteins can be produced via recombinant  
means in a variety of expression systems, e.g., various strains  
of bacterial, fungal, mammalian or insect cells. However, when  
bacteria are used as host cells for heterologous gene expression,  
several problems frequently occur.

20       For example, heterologous genes encoding small peptides are  
often poorly expressed in bacteria. Because of their size, most  
small peptides are unable to adopt stable, soluble conformations  
and are subject to intracellular degradation by proteases and  
peptidases present in the host cell. Those small peptides which  
25 do manage to accumulate when directly expressed in E. coli or  
other bacterial hosts are usually found in the insoluble or  
"inclusion body" fraction, an occurrence which renders them  
almost useless for screening purposes in biological or  
biochemical assays.

30       Moreover, even if small peptides are not produced in  
inclusion bodies, the production of small peptides by recombinant  
means as candidates for new drugs or enzyme inhibitors encounters  
further problems. Even small linear peptides can adopt an

enormous number of potential structures due to their degrees of conformational freedom. Thus a small peptide can have the 'desired' amino-acid sequence and yet have very low activity in an assay because the 'active' peptide conformation is only one of the many alternative structures adopted in free solution. This presents another difficulty encountered in producing small heterologous peptides recombinantly for effective research and therapeutic use.

Inclusion body formation is also frequently observed when the genes for heterologous proteins are expressed in bacterial cells. These inclusion bodies usually require further manipulations in order to solubilize and refold the heterologous protein, with conditions determined empirically and with uncertainty in each case.

If these additional procedures are not successful, little to no protein retaining bioactivity can be recovered from the host cells. Moreover, these additional processes are often technically difficult and prohibitively expensive for practical production of recombinant proteins for therapeutic, diagnostic or other research uses.

To overcome these problems, the art has employed certain peptides or proteins as fusion "partners" with a desired heterologous peptide or protein to enable the recombinant expression and/or secretion of small peptides or larger proteins as fusion proteins in bacterial expression systems. Among such fusion partners are included lacZ and trpE fusion proteins, maltose-binding protein fusions, and glutathione-S-transferase fusion proteins [See, generally, Current Protocols in Molecular Biology, Vol. 2, suppl. 10, publ. John Wiley and Sons, New York, NY, pp. 16.4.1-16.8.1 (1990); and Smith et al, Gene 67:31-40 (1988)]. U. S. Patent 4,801,536 describes the fusion of a bacterial flagellin protein to a desired protein to enable the production of a heterologous gene in a bacterial cell and its secretion into the culture medium as a fusion protein. PCT Patent Publication WO91/11454 discloses fusion proteins using

biotinylated renin as the fusion partner. The renin is immobilized on a purification column to facilitate separation and cleavage.

However, often fusions of desired peptides or proteins to other proteins (i.e., as fusion partners) at the amino- or carboxyl- termini of these fusion partner proteins have other potential disadvantages. Experience in E. coli has shown that a crucial factor in obtaining high levels of gene expression is the efficiency of translational initiation. Translational initiation in E. coli is very sensitive to the nucleotide sequence surrounding the initiating methionine codon of the desired heterologous peptide or protein sequence, although the rules governing this phenomenon are not clear. For this reason, fusions of sequences at the amino-terminus of many fusion partner proteins affects expression levels in an unpredictable manner. In addition there are numerous amino- and carboxy-peptidases in E. coli which degrade amino- or carboxyl-terminal peptide extensions to fusion partner proteins so that a number of the known fusion partners have a low success rate for producing stable fusion proteins.

The purification of proteins produced by recombinant expression systems is often a serious challenge. There is a continuing requirement for new and easier methods to produce homogeneous preparations of recombinant proteins, and yet a number of the fusion partners currently used in the art possess no inherent properties that would facilitate the purification process. Therefore, in the art of recombinant expression systems, there remains a need for new compositions and processes for the production and purification of stable, soluble peptides and proteins for use in research, diagnostic and therapeutic applications.

#### Summary of the Invention

In one aspect, the invention provides a fusion sequence comprising a thior doxin-like protein sequence fused to a

selected heterologous peptide or protein. The peptide or protein may be fused to the amino terminus of the thioredoxin-like sequence, the carboxyl terminus of the thioredoxin-like sequence, or within the thioredoxin-like sequence (e.g., within the active-site loop of thioredoxin). The fusion sequence according to this invention may optionally contain a linker peptide between the thioredoxin-like sequence and the selected peptide or protein. This linker provides, where needed, a selected cleavage site or a stretch of amino acids capable of preventing steric hindrance between the thioredoxin-like molecule and the selected peptide or protein.

As another aspect, the invention provides a DNA molecule encoding the fusion sequence defined above in association with, and under the control of, an expression control sequence capable of directing the expression of the fusion protein in a desired host cell.

Still a further aspect of the invention is a host cell transformed with, or having integrated into its genome, a DNA sequence comprising a thioredoxin-like DNA sequence fused to the DNA sequence of a selected heterologous peptide or protein. This fusion sequence is desirably under the control of an expression control sequence capable of directing the expression of a fusion protein in the cell.

As yet another aspect, there is provided a novel method for increasing the expression of soluble recombinant proteins. The method includes culturing under suitable conditions the above-described host cell to produce the fusion protein.

In one embodiment of this method, if the resulting fusion protein is cytoplasmic, the cell can be lysed by conventional means to obtain the soluble fusion protein. More preferably in the case of cytoplasmic fusion proteins, the method includes releasing the fusion protein from the host cell by applying osmotic shock or freeze/thaw treatments to the cell. In this case the fusion protein is selectively released from the interior of the cell via the zones of adhesion that exist between the

inner and outer membranes of E. coli. The fusion protein is then purified by conventional means.

In another embodiment of the method, if a secretory leader is employed in the fusion protein construct, the fusion protein  
5 can be recovered from a periplasmic extract or from the cell culture medium.

An additional step in both of these methods is cleavage of the desired protein from the thioredoxin-like protein by conventional means.

10 Other aspects and advantages of the present invention will be apparent upon consideration of the following detailed description of preferred embodiments thereof.

#### Summary of the Drawings

15 Fig. 1 illustrates the DNA sequence of the expression plasmid pALTRXA/EK/IL11 $\Delta$  Pro-581 and the amino acid sequence for the fusion protein therein, described in Example 1.

Fig. 2 illustrates the DNA sequence and amino acid sequence of the macrophage inhibitory protein-1 $\alpha$  (MIP-1 $\alpha$ ) protein used in  
20 the construction of a thioredoxin fusion protein described in Example 3.

Fig. 3 illustrates the DNA sequence and amino acid sequence of the bone morphogenetic protein-2 (BMP-2) protein used in the construction of a thioredoxin fusion protein described in Example  
25 4.

Fig. 4 is a schematic drawing illustrating the insertion of an enterokinase cleavage site into the active-site loop of E. coli thioredoxin (trxA) described in Example 5.

Fig. 5 is a schematic drawing illustrating random peptide  
30 insertions into the active-site loop of E. coli thioredoxin (trxA) described in Example 5.

Fig. 6 illustrates the DNA sequence and amino acid sequence of the human interleukin-6 (IL-6) protein used in the construction of a thioredoxin fusion protein described in Example  
35 6.

Fig. 7 illustrates the DNA sequence and amino acid sequence of the M-CSF protein used in the construction of a thioredoxin fusion protein described in Example 7.

5 Detailed Description of the Invention

This invention permits the production of large amounts of heterologous peptides or proteins in a stable, soluble form in certain host cells that normally express limited amounts of such peptides or proteins. It enables release of the fusion protein  
10 from the production cells without the necessity of lysing the cells, thereby streamlining the purification process. Also, by using a small peptide insert in an internal region of the thioredoxin-like sequence (e.g. the active site loop of thioredoxin) the invention provides a ready cleavage site,  
15 accessible on the surface of the molecule. The fusion proteins of this invention also permit the desired peptide or protein to achieve its desired conformation.

According to the present invention, the DNA sequence encoding a heterologous peptide or protein selected for  
20 expression in a recombinant system is fused to a thioredoxin-like DNA sequence for expression in the host cell. A thioredoxin-like DNA sequence is defined herein as a DNA sequence encoding a protein or fragment of a protein characterized by an amino acid sequence having at least 18% homology with the amino acid  
25 sequence of E. coli thioredoxin over an amino acid sequence length of 80 amino acids. Alternatively, a thioredoxin DNA sequence is defined as a DNA sequence encoding a protein or fragment of a protein characterized by a crystalline structure substantially similar to that of human or E. coli thioredoxin.  
30 The DNA sequence of glutaredoxin is one such sequence. The amino acid sequence of E. coli thioredoxin is described in H. Eklund et al, EMBO J. 3:1443-1449 (1984). The three-dimensional structure of E. coli thioredoxin is depicted in Fig. 2 of A. Holmgren, J. Biol. Chem. 264:13963-13966 (1989). Fig. 1 below nucleotide  
35 2242-2568 contains a DNA sequence encoding the E. coli



thioredoxin protein [Lim et al, J. Bacteriol. 163:311-316 (1985)]. The three latter publications are incorporated herein by reference for the purpose of providing information on thioredoxin which is known to one of skill in the art.

- 5       As the primary example of a thioredoxin-like protein useful in this invention, E. coli thioredoxin has the following characteristics. E. coli thioredoxin is a small protein, only 11.7 kD, and can be expressed to high levels (>10%, corresponding to a concentration of 15  $\mu$ M if cells are lysed at 10  $A_{550}/ml$ ).
- 10      The small size and capacity for high expression of the protein contributes to a high intracellular concentration. E. coli thioredoxin is further characterized by a very stable, tight structure which can minimize the effects on overall structural stability caused by fusion to the desired peptide or proteins.
- 15       The three dimensional structure of E. coli thioredoxin is known. It contains several surface loops, including a unique active site loop between residues Cys<sub>33</sub> and Cys<sub>36</sub> which protrudes from the body of the protein. This active site loop is an identifiable, accessible surface loop region and is not involved
- 20      in any interactions with the rest of the protein that contribute to overall structural stability. It is therefore a good candidate as a site for peptide insertions. Both the amino- and carboxyl-termini of E. coli thioredoxin are on the surface of the protein, and are readily accessible for fusions.
- 25       E. coli thioredoxin is also stable to proteases. Thus, E. coli thioredoxin may be desirable for use in E. coli expression systems, because as an E. coli protein it is characterized by stability to E. coli proteases. E. coli thioredoxin is also stable to heat up to 80°C and to low pH. Other thioredoxin-like
- 30      proteins encoded by thioredoxin-like DNA sequences useful in this invention may share the homologous amino acid sequences, and similar physical and structural characteristics. Thus, DNA sequences encoding other thioredoxin-like proteins may be used in place of E. coli thioredoxin according to this invention. For
- 35      example, the DNA sequence encoding other species' thioredoxin,

e.g., human thioredoxin, may be employed in the compositions and methods of this invention. Both the primary sequence and computer-predicted secondary structures of human and E. coli thioredoxins are very similar. Human thioredoxin also carries  
5 the same active site loop as is found in the E. coli protein. Insertions into the human thioredoxin active site loop and on the amino and carboxyl termini may be as well tolerated as those in E. coli thioredoxin.

Other thioredoxin-like sequences which may be employed in  
10 this invention include all or portions of the proteins glutaredoxin and various species' homologs thereof [A. Holmgren, cited above]. Although E. coli glutaredoxin and E. coli thioredoxin share less than 20% amino acid homology, the two proteins do have conformational and functional similarities  
15 [Eklund et al, EMBO J. 3:1443-1449 (1984)].

All or a portion of the DNA sequence encoding protein disulfide isomerase (PDI) and various species' homologs thereof [J. E. Edman et al, Nature 317:267-270 (1985)] may also be  
20 employed as a thioredoxin-like DNA sequence, since a repeated domain of PDI shares >18% homology with E. coli thioredoxin. The two latter publications are incorporated herein by reference for the purpose of providing information on glutaredoxin and PDI which is known and available to one of skill in the art.

Similarly the DNA sequence encoding phosphoinositide-  
25 specific phospholipase C (PI-PLC), fragments thereof and various species' homologs thereof [C. F. Bennett et al, Nature 334:268-270 (1988)] may also be employed in the present invention as a thioredoxin-like sequence based on the amino acid sequence homology with E. coli thioredoxin. All or a portion of the DNA  
30 sequence encoding an endoplasmic reticulum protein, such as Erp72, or various species homologs thereof are also included as thioredoxin-like DNA sequences for the purposes of this invention [R. A. Mazzarella et al, J. Biol. Chem. 265:1094-1101 (1990)] based on amino acid sequence homology. Another thioredoxin-like  
35 sequenc is a DNA sequenc which encodes all or a portion of an

adult T-cell leukemia-derived factor (ADF) or other species homologs thereof [N. Wakasugi et al, Proc. Natl. Acad. Sci., USA 87:8282-8286 (1990)] based on amino acid sequence homology to E. coli thioredoxin. The three latter publications are incorporated  
5 herein by reference for the purpose of providing information on PI-PLC, Erp72, and ADF which are known and available to one of skill in the art.

It is expected from the definition of thioredoxin-like DNA sequence used above that other sequences not specifically  
10 identified above, or perhaps not yet identified or published, may be useful as thioredoxin-like sequences based on their amino acid sequence similarities to E. coli thioredoxin and characteristic crystalline structural similarities to E. coli thioredoxin and the other thioredoxin-like proteins. Based on the above  
15 description, one of skill in the art should be able to select and identify, or, if desired, modify, a thioredoxin-like DNA sequence for use in this invention without resort to undue experimentation. For example, simple point mutations made to portions of native thioredoxin or native thioredoxin-like  
20 sequences which do not effect the structure of the resulting molecule are alternative thioredoxin-like sequences, as are allelic variants of native thioredoxin or native thioredoxin-like sequences.

DNA sequences which hybridize to the sequence for E. coli thioredoxin or its structural homologs under either stringent or  
25 relaxed hybridization also encode thioredoxin-like proteins for use in this invention. Stringent hybridization is defined herein as hybridization at 4XSSC at 65°C, followed by a washing in 0.1XSSC at 65°C for an hour. Alternatively stringent  
30 hybridization is defined as hybridization in 50% formamide, 4XSSC at 42°C. Non-stringent hybridization is defined herein as hybridizing in at are 4XSSC at 50°C, or hybridization with 30-40% formamide at 42°C. The use of all such thioredoxin-like sequences are believed to be encompassed in this invention.

35 C nstructi n f a fusi n s quenc f th present invention,

which comprises the DNA sequence of a selected peptide or protein and the DNA sequence of a thioredoxin-like sequence, employs conventional genetic engineering techniques [see, Sambrook et al, Molecular Cloning. A Laboratory Manual., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1989)]. Fusion sequences may be prepared in a number of different ways. For example, the selected heterologous protein may be fused to the amino terminus of the thioredoxin-like molecule. Alternatively, the selected protein sequence may be fused to the carboxyl terminus of the thioredoxin-like molecule. Small peptide sequences could also be fused to either of the above-mentioned positions of the thioredoxin-like sequence to produce them in a structurally unconstrained manner.

This fusion of a desired heterologous peptide or protein to the thioredoxin-like protein increases the stability of the peptide or protein. At either the amino or carboxyl terminus, the desired heterologous peptide or protein is fused in such a manner that the fusion does not destabilize the native structure of either protein. Additionally, fusion to the soluble thioredoxin-like protein improves the solubility of the selected heterologous peptide or protein.

It may be preferred for a variety of reasons that peptides be fused within the active site loop of the thioredoxin-like molecule. The face of thioredoxin surrounding the active site loop has evolved, in keeping with the protein's major function as a nonspecific protein disulfide oxido-reductase, to be able to interact with a wide variety of protein surfaces. The active site loop region is found between segments of strong secondary structure and offers many advantages for peptide fusions. A small peptide inserted into the active-site loop of a thioredoxin-like protein is present in a region of the protein which is not involved in maintaining tertiary structure. Therefore the structure of such a fusion protein should be stable. Previous work has shown that E. coli thioredoxin can be cleaved into two fragments at a position close to the active site

loop, and yet the tertiary interactions stabilizing the protein remain.

The active site loop of E. coli thioredoxin has the sequence NH<sub>2</sub>...Cys<sub>33</sub>-Gly-Pro-Cys<sub>36</sub>...COOH. Fusing a selected peptide with  
5 a thioredoxin-like protein in the active loop portion of the protein constrains the peptide at both ends, reducing the degrees of conformational freedom of the peptide, and consequently reducing the number of alternative structures taken by the peptide. The inserted peptide is bound at each end by cysteine  
10 residues, which may form a disulfide linkage to each other as they do in native thioredoxin and further limit the conformational freedom of the inserted peptide.

Moreover, this invention places the peptide on the surface of the thioredoxin-like protein. Thus the invention provides a  
15 distinct advantage for use of the peptides in screening for bioactive peptide conformations and other assays by presenting peptides inserted in the active site loop in this structural context.

Additionally the fusion of a peptide into the loop protects  
20 it from the actions of E. coli amino- and carboxyl-peptidases. Further a restriction endonuclease cleavage site RsrII already exists in the portion of the E. coli thioredoxin DNA sequence encoding the loop region at precisely the correct position for a peptide fusion [see Figure 4]. RsrII recognizes the DNA sequence  
25 CGG(A/T)CCG leaving a three nucleotide long 5'-protruding sticky end. DNA bearing the complementary sticky ends will therefore insert at this site in just one orientation.

A fusion sequence of a thioredoxin-like sequence and a desired protein or peptide sequence according to this invention  
30 may optionally contain a linker peptide inserted between the thioredoxin-like sequence and the selected heterologous peptide or protein. This linker sequence may encode, if desired, a polypeptide which is selectably cleavable or digestible by conventional chemical or enzymatic methods. For example, the  
35 selected cleavage site may be an enzymatic cleavage site.

Examples of enzymatic cleavage sites include sites for cleavage by a proteolytic enzyme, such as enterokinase, Factor Xa, trypsin, collagenase, and thrombin. Alternatively, the cleavage site in the linker may be a site capable of being cleaved upon exposure to a selected chemical, e.g., cyanogen bromide, hydroxylamine, or low pH.

Cleavage at the selected cleavage site enables separation of the heterologous protein or peptide from the thioredoxin fusion protein to yield the mature heterologous peptide or protein. The mature peptide or protein may then be obtained in purified form, free from any polypeptide fragment of the thioredoxin-like protein to which it was previously linked. The cleavage site, if inserted into a linker useful in the fusion sequences of this invention, does not limit this invention. Any desired cleavage site, of which many are known in the art, may be used for this purpose.

The optional linker sequence of a fusion sequence of the present invention may serve a purpose other than the provision of a cleavage site. The linker may also be a simple amino acid sequence of a sufficient length to prevent any steric hindrance between the thioredoxin-like molecule and the selected heterologous peptide or protein.

Whether or not such a linker sequence is necessary will depend upon the structural characteristics of the selected heterologous peptide or protein and whether or not the resulting fusion protein is useful without cleavage. For example, where the thioredoxin-like sequence is a human sequence, the fusion protein may itself be useful as a therapeutic without cleavage of the selected protein or peptide therefrom. Alternatively, where the mature protein sequence may be naturally cleaved, no linker may be needed.

In one embodiment therefore, the fusion sequence of this invention contains a thioredoxin-like sequence fused directly at its amino or carboxyl terminal end to the sequence of the selected polypeptide or protein. The resulting fusion protein is

thus a soluble cytoplasmic fusion protein. In another embodiment, the fusion sequence further comprises a linker sequence interposed between the thioredoxin-like sequence and the selected peptide or protein sequence. This fusion protein is  
5 also produced as a soluble cytoplasmic protein. Similarly, where the selected peptide sequence is inserted into the active site loop region or elsewhere within the thioredoxin-like sequence, a cytoplasmic fusion protein is produced.

The cytoplasmic fusion protein can be purified by  
10 conventional means. Preferably, as a novel aspect of the present invention, several thioredoxin fusion proteins of this invention may be purified by exploiting an unusual property of thioredoxin. The cytoplasm of E. coli is effectively isolated from the external medium by a cell envelope comprising two membranes,  
15 inner and outer, separated from each other by a periplasmic space within which lies a rigid peptidoglycan cell wall. The peptidoglycan wall contributes both shape and strength to the cell. At certain locations in the cell envelope there are "gaps" (called variously Bayer patches, Bayer junctions or adhesion  
20 sites) in the peptidoglycan wall where the inner and outer membranes appear to meet and perhaps fuse together. See, M. E. Bayer, J. Bacteriol. 93:1104-1112 (1967) and J. Gen. Microbiol. 53:395-404 (1968). Most of the cellular thioredoxin lies loosely associated with the inner surface of the membrane at these  
25 adhesion sites and can be quantitatively expelled from the cell through these adhesion sites by a sudden osmotic shock or by a simple freeze/thaw procedure. See C. A. Lunn and V. P. Pigiet, J. Biol. Chem. 257:11424-11430 (1982) and in "Thioredoxin and Glutaredoxin Systems: Structure and Function:165-176 (1986) ed.  
30 A. Holmgren et al., Raven Press, New York. To a lesser extent some EF-Tu (elongation factor-Tu) can be expelled in the same way [Jacobson et al, Biochemistry 15:2297-2302 (1976)], but, with the exception of the periplasmic contents, the vast majority of E. coli proteins cannot be released by these treatments.

35 Although there have been reports of the release by osmotic

shock of a limited number of heterologous proteins produced in the cytoplasm of E. coli [Denefle et al, Gene 85:499-510 (1989); Joseph-Liauzun et al, Gene 86:291-295 (1990); Rosenwasser et al, J. Biol. Chem. 265:13066-13073 (1990)], the ability to be so released is a rare and desirable property not shared by the majority of heterologous proteins. Fusion of a heterologous protein to thioredoxin as described by the present invention not only enhances its expression, solubility and stability as described above, but may also provide for its release from the cell by osmotic shock or freeze/thaw treatments, greatly simplifying its purification. The thioredoxin portion of the fusion protein in some cases, e.g., with MIP, directs the fusion protein towards the adhesion sites, from where it can be released to the exterior by these treatments.

In another embodiment the present invention may employ another component, that is, a secretory leader sequence, among which many are known in the art, e.g. leader sequences of phoA, MBP,  $\beta$ -lactamase, operatively linked in frame to the fusion protein of this invention to enable the expression and secretion of the mature fusion protein into the bacterial periplasmic space or culture medium. This leader sequence may be fused to the amino terminus of the thioredoxin-like molecule when the selected peptide or protein sequence is fused to the carboxyl terminus or to an internal site within the thioredoxin-like sequence. An optional linker could also be present when the peptide or protein is fused at the carboxyl terminus. It is expected that this fusion sequence construct when expressed in an appropriate host cell would be expressed as a secreted fusion protein rather than a cytoplasmic fusion protein. However stability, solubility and high expression should characterize fusion proteins produced using any of these alternative embodiments.

This invention is not limited to any specific type of heterologous peptide or protein. A wide variety of heterologous genes or gene fragments are useful in forming the fusion sequences of the present invention. While the compositions and



methods of this invention are most useful for peptides or proteins which are not expressed, expressed in inclusion bodies, or expressed in very small amounts in bacterial and yeast hosts, the heterologous peptides or proteins can include any peptide or protein useful for human or veterinary therapy, diagnostic or research applications in any expression system. For example, hormones, cytokines, growth or inhibitory factors, enzymes, modified or wholly synthetic proteins or peptides can be produced according to this invention in bacterial, yeast, mammalian or other eukaryotic cells and expression systems suitable therefor.

In the examples below illustrating this invention, the proteins expressed by this invention include IL-11, MIP-1 $\alpha$ , IL-6, M-CSF, a bone inductive factor called BMP-2, and a variety of small peptides of random sequence. These proteins include examples of proteins which, when expressed without a thioredoxin fusion partner, are unstable in E. coli or are found in inclusion bodies.

A variety of DNA molecules incorporating the above-described fusion sequences may be constructed for expressing the heterologous peptide or protein according to this invention. At a minimum a desirable DNA sequence according to this invention comprises a fusion sequence described above, in association with, and under the control of, an expression control sequence capable of directing the expression of the fusion protein in a desired host cell. For example, where the host cell is an E. coli strain, the DNA molecule desirably contains a promoter which functions in E. coli, a ribosome binding site, and optionally, a selectable marker gene and an origin of replication if the DNA molecule is extra chromosomal. Numerous bacterial expression vectors containing these components are known in the art for bacterial expression, and can easily be constructed by standard molecular biology techniques. Similarly known yeast and mammalian cell vectors and vector components may be utilized where the host cell is a yeast cell or a mammalian cell.

The DNA molecules containing the fusion sequences may be

further modified to contain different codons to optimize expression in the selected host cell, as is known in the art.

These DNA molecules may additionally contain multiple copies of the thioredoxin-like DNA sequence, with the heterologous protein fused to only one of the DNA sequences, or with the heterologous protein fused to all copies of the thioredoxin-like sequence. It may also be possible to integrate a thioredoxin-like/heterologous peptide or protein-encoding fusion sequence into the chromosome of a selected host to either replace or duplicate a native thioredoxin-like sequence.

Host cells suitable for the present invention are preferably bacterial cells. For example, the various strains of E. coli (e.g., HB101, W3110 and strains used in the following examples) are well-known as host cells in the field of biotechnology. E. coli strain GI724, used in the following examples, has been deposited with a United States microorganism depository as described in detail below. Various strains of B. subtilis, Pseudomonas, and other bacteria may also be employed in this method.

Many strains of yeast and other eukaryotic cells known to those skilled in the art may also be useful as host cells for expression of the polypeptides of the present invention. Similarly known mammalian cells may also be employed in the expression of these fusion proteins.

To produce the fusion protein of this invention, the host cell is either transformed with, or has integrated into its genome, a DNA molecule comprising a thioredoxin-like DNA sequence fused to the DNA sequence of a selected heterologous peptide or protein, desirably under the control of an expression control sequence capable of directing the expression of a fusion protein. The host cell is then cultured under known conditions suitable for fusion protein production. If the fusion protein accumulates in the cytoplasm of the cell it may be released by conventional bacterial cell lysis techniques and purified by conventional procedures including selective precipitations, solubilizations

and column chromatographic methods. If a secretory leader is incorporated into the fusion molecule substantial purification is achieved when the fusion protein is secreted into the periplasmic space or the growth medium.

5       Alternatively, for cytoplasmic thioredoxin fusion proteins, a selective release from the cell may be achieved by osmotic shock or freeze/thaw procedures. Although final purification is still required for most purposes, the initial purity of fusion proteins in preparations resulting from these procedures is  
10 superior to that obtained in conventional whole cell lysates, reducing the number of subsequent purification steps required to attain homogeneity. In a typical osmotic shock procedure, the packed cells containing the fusion protein are resuspended on ice in a buffer containing EDTA and having a high osmolarity, usually  
15 due to the inclusion of a solute, such as 20% w/v sucrose, in the buffer which cannot readily cross the cytoplasmic membrane. During a brief incubation on ice the cells plasmolyze as water leaves the cytoplasm down the osmotic gradient. The cells are then switched into a buffer of low osmolarity, and during the  
20 osmotic re-equilibration both the contents of the periplasm and proteins localized at the Bayer patches are released to the exterior. A simple centrifugation following this release removes the majority of bacterial cell-derived contaminants from the fusion protein preparation. Alternatively, in a freeze/thaw  
25 procedure the packed cells containing the fusion protein are first resuspended in a buffer containing EDTA and are then frozen. Fusion protein release is subsequently achieved by allowing the frozen cell suspension to thaw. The majority of contaminants can be removed as described above by a  
30 centrifugation step. The fusion protein is further purified by well-known conventional methods.

These treatments typically release at least 30% of the fusion proteins without lysing the cell cultures. The success of these procedures in releasing significant amounts of a wide  
35 variety of thioredoxin fusion proteins is surprising, since such

techniques are not generally successful with a wide range of proteins. The ability of these fusion proteins to be substantially purified by such treatments, which are significantly simpler and less expensive than the purification methods required by other fusion protein systems, may provide the fusion proteins of the invention with a significant advantage over other systems which are used to produce proteins in E. coli.

The resulting fusion protein is stable and soluble, often with the heterologous peptide or protein retaining its bioactivity. The heterologous peptide or protein may optionally be separated from the thioredoxin-like protein by cleavage, as discussed above.

In the specific and illustrative embodiments of the compositions and methods of this invention, the E. coli thioredoxin (trxA) gene has been cloned and placed in an E. coli expression system. An expression plasmid pALtrxA-781 was constructed. This plasmid containing modified IL-11 fused to the thioredoxin sequence and called pALtrxA/EK/IL11 Pro-581 is described below in Example 1 and in Fig. 1. A modified version of this plasmid containing a different ribosome binding site was employed in the other examples and is specifically described in Example 3. Other conventional vectors may be employed in this invention. The invention is not limited to the plasmids described in these examples.

Plasmid pALtrxA-781 (without the modified IL-11) directs the accumulation of >10% of the total cell protein as thioredoxin in E. coli host strain GI724. Examples 2 through 6 describe the use of this plasmid to form and express thioredoxin fusion proteins with BMP-2, IL-6 and MIP-1 $\alpha$ , which are polypeptides.

As an example of the expression of small peptides inserted into the active-site loop, a derivative of pALtrxA-781 has been constructed in which a 13 amino-acid linker peptide sequence containing a cleavage site for the specific protease enterokinase [Leipnieks and Light, J. Biol. Chem. 254:1077-1083 (1979)] has been fused into the active site loop of thioredoxin. This

plasmid (pALtrxA-EK) directs the accumulation of >10% of the total cell protein as the fusion protein. The fusion protein is all soluble, indicating that it has probably adopted a 'native' tertiary structure. It is equally as stable as wild type

5 thiorodoxin to prolonged incubations at 80°C, suggesting that the strong tertiary structure of thiorodoxin has not been compromised by the insertion into the active site loop. The fusion protein is specifically cleaved by enterokinase, whereas thiorodoxin is not, indicating that the peptide inserted into the active site

10 loop is present on the surface of the fusion protein.

As described in more detail in Example 5 below, fusions of small peptides were made into the active site loop of thiorodoxin. The inserted peptides were 14 residues long and were of totally random composition to test the ability of the

15 system to deal with hydrophobic, hydrophilic and neutral sequences.

The methods and compositions of this invention permit the production of proteins and peptides useful in research, diagnostic and therapeutic fields. The production of fusion

20 proteins according to this invention has a number of advantages. As one example, the production of a selected protein by the present invention as a carboxyl-terminal fusion to E. coli thiorodoxin, or another thiorodoxin-like protein, enables avoidance of translation initiation problems often encountered in

25 the production of eukaryotic proteins in E. coli. Additionally the initiator methionine usually remaining on the amino-terminus of the heterologous protein is not present and does not have to be removed when the heterologous protein is made as a carboxyl terminal thiorodoxin fusion.

30 The production of fusion proteins according to this invention reliably improves solubility of desired heterologous proteins and enhances their stability to proteases in the expression system. This invention also enables high level

35 xpr ssi n f certain desirable th rapeutic prot ins, e.g., IL-11, which are otherwise produced at low levels in bact rial host

cells.

This invention may also confer heat stability to the fusion protein, especially if the heterologous protein itself is heat stable. Because thioredoxin, and presumably all thioredoxin-like  
5 proteins, are heat stable up to 80°C, the present invention may enable the use of a simple heat treatment as an initial effective purification step for some thioredoxin fusion proteins.

In addition to providing high levels of the selected heterologous proteins or peptides upon cleavage from the fusion  
10 protein for therapeutic or other uses, the fusion proteins or fusion peptides of the present invention may themselves be useful as therapeutics. Further the thioredoxin-like fusion proteins may provide a vehicle for the delivery of bioactive peptides. As one example, human thioredoxin would not be antigenic in humans,  
15 and therefore a fusion protein of the present invention with human thioredoxin may be useful as a vehicle for delivering to humans the biologically active peptide to which it is fused. Because human thioredoxin is an intracellular protein, human thioredoxin fusion proteins may be produced in an E. coli  
20 intracellular expression system. Thus this invention also provides a method for delivering biologically active peptides or proteins to a patient in the form of a fusion protein with an acceptable thioredoxin-like protein.

The present invention also provides methods and reagents for  
25 screening libraries of random peptides for their potential enzyme inhibitory, hormone/growth factor agonist and hormone/growth factor antagonist activity. Also provided are methods and reagents for the mapping of known protein sequences for regions of potential interest, including receptor binding sites,  
30 substrate binding sites, phosphorylation/modification sites, protease cleavage sites, and epitopes.

Bacterial colonies expressing thioredoxin-like/random peptide fusion proteins may be screened using radiolabelled  
prot ins such as hormones r growth fact rs as prob s. Positives  
35 arising from this typ of screen w uld identify mimics of

receptor binding sites and may lead to the design of compounds with therapeutic uses. Bacterial colonies expressing thioredoxin-like random peptide fusion proteins may also be screened using antibodies raised against native, active hormones or growth factors. Positives arising from this type of screen could be mimics of surface epitopes present on the original antigen. Where such surface epitopes are responsible for receptor binding, the 'positive' fusion proteins would have biological activity.

Additionally, the thioredoxin-like fusion proteins or fusion peptides of this invention may also be employed to develop monoclonal and polyclonal antibodies, or recombinant antibodies or chimeric antibodies, generated by known methods for diagnostic, purification or therapeutic use. Studies of thioredoxin-like molecules indicate a possible B cell/T cell growth factor activity [N. Wakasuki et al, cited above], which may enhance immune response. The fusion proteins or peptides of the present invention may be employed as antigens to elicit desirable antibodies, which themselves may be further manipulated by known techniques into monoclonal or recombinant antibodies.

Alternatively, antibodies elicited to thioredoxin-like sequences may also be useful in the purification of many different thioredoxin fusion proteins.

The following examples illustrate embodiments of the present invention, but are not intended to limit the scope of the disclosure.

#### EXAMPLE 1 - THIOREDOXIN-IL-11 FUSION MOLECULE

A thioredoxin-like fusion molecule of the present invention was prepared using E. coli thioredoxin as the thioredoxin-like sequence and recombinant IL-11 as the selected heterologous protein. The DNA and amino acid sequence of IL-11 has been published. See Paul et al, Proc. Natl. Acad. Sci. U.S.A. 87:7512-7516 (1990) and PCT Patent publication WO91/0749, published May 30, 1991. IL-11 DNA can be obtained by cloning

based on its published sequence. The E. coli thioredoxin (trxA) gene was cloned based on its published sequence and employed to construct various related E. coli expression plasmids using standard DNA manipulation techniques, described extensively by  
5 Sambrook, Fritsch and Maniatis, Molecular Cloning. A Laboratory Manual, 2nd edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989).

A first expression plasmid pALTRxa-781 was constructed containing the E. coli trxA gene without fusion to another  
10 sequence. This plasmid further contained sequences which are described in detail below for the related IL-11 fusion plasmid. This first plasmid, which directs the accumulation of >10% of the total cell protein as thioredoxin in an E. coli host strain GI724, was further manipulated as described below for the  
15 construction of a trxA/IL-11 fusion sequence.

The entire sequence of the related plasmid expression vector, pALtrxA/EK/IL11a Pro-581, is illustrated in Fig. 1 and contains the following principal features:

Nucleotides 1-2060 contain DNA sequences originating from  
20 the plasmid pUC-18 [Norrande et al, Gene 26: 101-106 (1983)] including sequences containing the gene for  $\beta$ -lactamase which confers resistance to the antibiotic ampicillin in host E. coli strains, and a colE1-derived origin of replication. Nucleotides 2061-2221 contain DNA sequences for the major leftward promoter  
25 (pL) of bacteriophage  $\lambda$  [Sanger et al, J. Mol. Biol. 162:729-773 (1982)], including three operator sequences, O<sub>L</sub>1, O<sub>L</sub>2 and O<sub>L</sub>3. The operators are the binding sites for  $\lambda$  cI repressor protein, intracellular levels of which control the amount of transcription initiation from pL. Nucleotides 2222-2241 contain a strong  
30 ribosome binding sequence derived from that of gene 10 of bacteriophage T7 [Dunn and Studier J. Mol. Biol. 166:477-535 (1983)].

Nucleotides 2242-2568 contain a DNA sequence encoding the E. coli thioredoxin protein [Lim et al, J. Bacteriol. 163:311-316  
35 (1985)]. There is no translation termination codon at the end of



the thioredoxin coding sequence in this plasmid.

Nucleotides 2569-2583 contain DNA sequence encoding the amino acid sequence for a short, hydrophilic, flexible spacer peptide "--GSGSG--". Nucleotides 2584-2598 provide DNA sequence  
5 encoding the amino acid sequence for the cleavage recognition site of enterokinase (EC 3.4.4.8), "--DDDDK--" [Maroux et al, J. Biol. Chem. 246:5031-5039 (1971)].

Nucleotides 2599-3132 contain DNA sequence encoding the amino acid sequence of a modified form of mature human IL-11  
10 [Paul et al, Proc. Natl. Acad. Sci. USA 87:7512-7516 (1990)], deleted for the N-terminal prolyl-residue normally found in the natural protein. The sequence includes a translation termination codon at the 3'-end of the IL-11 sequence.

Nucleotides 3133-3159 provide a "Linker" DNA sequence  
15 containing restriction endonuclease sites. Nucleotides 3160-3232 provide a transcription termination sequence based on that of the E. coli aspA gene [Takagi et al, Nucl. Acids Res. 13:2063-2074 (1985)]. Nucleotides 3233-3632 are DNA sequences derived from pUC-18.

20 As described in Example 2 below, when cultured under the appropriate conditions in a suitable E. coli host strain, this plasmid vector can direct the production of high levels (approximately 10% of the total cellular protein) of a thioredoxin-IL-11 fusion protein. By contrast, when not fused to  
25 thioredoxin, IL-11 accumulated to only 0.2% of the total cellular protein when expressed in an analogous host/vector system.

#### EXAMPLE 2 - EXPRESSION OF A FUSION PROTEIN

A thioredoxin-IL-11 fusion protein was produced according to  
30 the following protocol using the plasmid constructed as described in Example 1. pALtrxA/EK/IL11Δ Pro-581 was transformed into the E. coli host strain GI724 (F<sup>-</sup>, lacI<sup>q</sup>, lacP<sup>L8</sup>, amp<sup>c</sup>:Δ cI<sup>+</sup>) by the procedure of Dagert and Ehrlich, Gene 6:23 (1979). The untransformed host strain E. coli GI724 was deposited with the  
35 American Type Culture Collection, 12301 Parklawn Drive ,

Rockville, Maryland on January 31, 1991 under ATCC No. 55151 for patent purposes pursuant to applicable laws and regulations. Transformants were selected on 1.5% w/v agar plates containing IMC medium, which is composed of M9 medium [Miller, "Experiments in Molecular Genetics", Cold Spring Harbor Laboratory, New York (1972)] supplemented with 0.5% w/v glucose, 0.2% w/v casamino acids and 100 µg/ml ampicillin.

GI724 contains a copy of the wild-type  $\lambda$  cI repressor gene stably integrated into the chromosome at the ampC locus, where it has been placed under the transcriptional control of Salmonella typhimurium trp promoter/operator sequences. In GI724,  $\lambda$  cI protein is made only during growth in tryptophan-free media, such as minimal media or a minimal medium supplemented with casamino acids such as IMC, described above. Addition of tryptophan to a culture of GI724 will repress the trp promoter and turn off synthesis of  $\lambda$  cI, gradually causing the induction of transcription from pL promoters if they are present in the cell.

GI724 transformed with pALtrxA/EK/IL11A Pro-581 was grown at 37°C to an A<sub>550</sub> of 0.5 in IMC medium. Tryptophan was added to a final concentration of 100 µg/ml and the culture incubated for a further 4 hours. During this time thioredoxin-IL-11 fusion protein accumulated to approximately 10% of the total cell protein.

All of the fusion protein was found to be in the soluble cellular fraction, and was purified as follows. Cells were lysed in a french pressure cell at 20,000 psi in 50 mM HEPES pH 8.0, 1 mM phenylmethylsulfonyl fluoride. The lysate was clarified by centrifugation at 15,000 x g for 30 minutes and the supernatant loaded onto a QAE-Toyopearl column. The flow-through fractions were discarded and the fusion protein eluted with 50 mM HEPES pH 8.0, 100 mM NaCl. The eluate was adjusted to 2M NaCl and loaded onto a column of phenyl-Toyopearl. The flow-through fractions were again discarded and the fusion protein eluted with 50 mM HEPES pH 8.0, 0.5 M NaCl.

The fusion protein was then dialyzed against 25 mM HEPES pH

8.0 and was >80% pure at this stage. By T1165 bioassay [Paul et al, cited above] the purified thioredoxin-IL-11 protein exhibited an activity of  $8 \times 10^5$  U/mg. This value agrees closely on a molar basis with the activity of  $2 \times 10^6$  U/mg found for COS cell-derived IL-11 purified to homogeneity and measured for activity in the same assay. One milligram of the fusion protein was then cleaved at 37°C for 20 hours with 1000 units of bovine enterokinase [Leipnieks and Light, *J. Biol. Chem.* 254:1677-1683 (1979)] in 1 ml 10mM Tris-Cl (pH 8.0)/10mM  $\text{CaCl}_2$ . IL-11 was recovered from the reaction products by passing them over a QAE-Toyopearl column in 25 mM HEPES pH 8.0, where homogeneous IL-11 was found in the flow-through fractions. Uncleaved fusion protein, thioredoxin and enterokinase remained bound on the column.

The homogeneous IL-11 prepared in this manner had a bioactivity in the T1165 assay of  $2.5 \times 10^6$  U/mg. Its physical and chemical properties were determined as follows:

(1) Molecular Weight

The molecular weight of the IL-11 was found to be about 21 kD as measured by 10% SDS-PAGE under reducing conditions (tricine system) in accordance with the methods of Schagger, et al., *Anal Biochem.* 166:368-379. (1987). The compound ran as a single band.

(2) Endotoxin Content

The endotoxin content of the IL-11 was found to be less than 0.1 nanogram per milligram IL-11 in the LAL (*Limulus* amoebocyte lysate, Pyrotel, available from Associates of Cape Cod, Inc., Woods Hole, Massachusetts, U.S.A.) assay, conducted in accordance with the manufacturer's instructions.

(3) Isoelectric Point

The theoretical isoelectric point of IL-11 is pH 11.70. As measured by polyacrylamide gel isoelectric focusing using an LKB Ampholine PAGplate with a pH range from 3.5 to 9.5, the IL-11 ran at greater than 9.5. An exact measurement could not be taken

because IL-11 is too basic a protein for the reliable gels available.

(4) Fluorescence Absorption Spectrum

5 Fluorescence absorption spectrum of the IL-11, as measured on a 0.1% aqueous solution in a 1 cm quartz cell showed an emission maximum at 335-337 nm.

(5) UV Absorption

10 UV absorption of the IL-11 on a 0.1% aqueous solution in a 1 cm quartz cell showed an absorbance maximum at 278-280 nm.

(6) Amino Acid Composition

15 The theoretical amino acid composition for IL-11, based on its amino acid sequence is as follow:

	<u>Amino Acid</u>	<u>Number</u>	<u>Mole %</u>
	Ala	20	11.3
	Asp Acid	11	6.22
20	Cysteine	0	
	Glu	3	1.70
	Phe	1	0.57
	Gly	14	7.91
	His	4	2.26
25	Ile	2	1.13
	Lys	3	1.70
	Leu	41	23.16
	Met	2	1.13
	Asn	1	0.57
30	Pro	21	11.86
	Gln	7	3.96
	Arg	18	10.17
	Ser	11	6.22
	Thr	9	5.09
35	Val	5	2.83
	Trp	3	1.70
	Tyr	1	0.57

A sample of homogenous IL-11 was subjected to vapor phase  
40 hydrolysis as follows:

6 N HCl and 2 N Ph nol reagent wer added to hydrolysis vessel in which tubes containing 45 µl f 1:10 dilut d (w/H<sub>2</sub>O)

IL-11, concentrated to dryness are inserted. Samples were sealed under vacuum and hydrolyzed for 36 hours at 110°C. After the hydrolysis, samples were dried and resuspended in 500 µl Na-S sample dilution buffer. Amino acid analysis was performed on a Beckman 7300 automated amino acid analyzer. A cation exchange column was used for separation of amino acids following post column derivatization with ninhydrin. Primary amino acids were detected at 570 nm and secondary amino acids were detected at 440 nm. Eight point calibration curves were constructed for each of the amino acids.

Because certain amino acids are typically not recovered, results for only 5 amino acids are given below. Since the hydrolysis was done without desalting the protein, 100% recovery was achieved for most of the amino acids.

The relative recovery of each individual amino acid residue per molecule of recombinant IL-11 was determined by normalizing GLX = 10 (the predicted number of glutamine and glutamic acid residue in IL-11 based on cDNA sequence). The value obtained for the recovery of GLX in picomoles was divided by 10 to obtain the GLX quotient. Dividing the value obtained for the recovery in picomoles of each amino acid by the GLX quotient for that sample gives a number that represents the relative recovery of each amino acid in the sample, normalized to the quantitative recovery of GLX residues. The correlation coefficient comparing the expected versus the average number of residues of each amino acid observed is greater than 0.985, indicating that the number of residues observed for each amino acid is in good agreement with that predicted sequence.

30	0	Amino	1	No. of Residues	2	No. of Residues	3	Correlation
		Acids		Calculated		Expected		Coefficient
	1	Asp		12.78		12		
	2	Glu		10.00		10		
	3	Gly		12.80		14		0.9852
35	4	Arg		16.10		18		
	5	Pro		18.40		21		

(7) Amino Terminus Sequencing

IL-11 (buffered in 95% acetonitrile TFA) was sequenced using an ABI 471A protein sequencer (ABI, Inc.) in accordance with the manufacturer's instructions. Amino terminus sequencing confirmed  
5 that the thioredoxin fusion protein produced IL-11 contained the correct IL-11 amino acid sequence and only one amino terminus observed.

(8) Peptide Mapping

10 The IL-11 was cleaved with Endoproteinase Asp-N (Boehringer Mannheim) (1:500 ratio of Asp-N to IL-11) in 10 mM Tris, pH 8, 1 M urea and 2 mM 4-aminobenzamidine dihydrochloride (PABA), at 37°C for 4 hours. The sample was then run on HPLC on a C4 Vydac column using an A buffer of 50 mM NaHPO<sub>4</sub>, pH 4.3, in dH<sub>2</sub>O, a B  
15 buffer of 100% isopropanol with a gradient at 1 ml/min from 100%A to 25%A and 75%B (changing 1%/minute). The eluted peptide fragments were then sequenced using an ABI 471A protein sequencer (ABI, Inc.) in accordance with the manufacturer's instructions. Peptide mapped confirmed the IL-11 produced from the thioredoxin  
20 fusion protein contained the proper IL-11 N-terminal and C-terminal sequences.

(9) Solubility

IL-11 protein was tested for solubility in the substances  
25 below with the following results:

Water	very soluble
Ethyl Alcohol	very soluble
Acetone	very soluble
1M sodium chloride	very soluble
30 10% sucrose	very soluble

(10) Sugar Composition and Protein/Polysaccharide Content in %

The absence of sugar moieties attached to the polypeptide  
35 backbone of the IL-11 protein is indicated by its amino acid sequenc , which contains none f the typical sugar attachment sit s.

**EXAMPLE 3 - THIOREDOXIN-MIP FUSION MOLECULE**

Human macrophage inflammatory protein 1 $\alpha$  (MIP-1 $\alpha$ ) was expressed at high levels in *E. coli* as a thioredoxin fusion protein using an expression vector similar to pALtrxA/EK/IL11 $\alpha$  Pro-581 described in Example 1 above but modified in the following manner to replace the ribosome binding site of bacteriophage T7 with that of  $\lambda$ CII. In the plasmid of Example 1, nucleotides 2222 and 2241 were removed by conventional means. Inserted in place of those nucleotides was a sequence of nucleotides formed by nucleotides 35566 to 35472 and 38137 to 38361 from bacteriophage lambda as described in Sanger et al (1982) cited above. This reference is incorporated by reference for the purpose of disclosing this sequence. To express a thioredoxin-MIP-1 $\alpha$  fusion the DNA sequence in the thusly-modified pALtrxA/EK/IL11 $\alpha$  Pro-581 encoding human IL-11 (nucleotides 2599-3132) is replaced by the 213 nucleotide DNA sequence shown in Fig. 2 encoding full-length, mature human MIP-1 $\alpha$  [Nakao et al, *Mol. Cell. Biol.* 10:3646-3658 (1990)].

The host strain and expression protocol used for the production of thioredoxin-MIP-1 $\alpha$  fusion protein are as described in Example 1. As was seen with the thioredoxin-IL-11 fusion protein, all of the thioredoxin-MIP-1 $\alpha$  fusion protein was found in the soluble cellular fraction, representing up to 20% of the total protein.

Cells were lysed as in Example 1 to give a protein concentration in the crude lysate of 10 mg/ml. This lysate was then heated at 80°C for 10 min to precipitate the majority of contaminating *E. coli* proteins and was clarified by centrifugation at 130,000 x g for 60 minutes. The pellet was discarded and the supernatant loaded onto a Mono Q column. The fusion protein eluted at approximately 0.5 M NaCl from this column and was >80% pure at this stage. After dialysis to remove salt the fusion protein could be cleaved by an enterokinase treatment as described in Example 1 to r 1 as MIP-1 $\alpha$ .

**EXAMPLE 4 - THIOREDOXIN-BMP-2 FUSION MOLECULE**

Human Bone Morphogenetic Protein 2 (BMP-2) was expressed at high levels in E. coli as a thioredoxin fusion protein using the modified expression vector described in Example 3. The DNA sequence encoding human IL-11 in the modified pALtrxA/EK/IL11Δ Pro-581 (nucleotides 2599-3132) is replaced by the 345 nucleotide DNA sequence shown in Fig. 3 encoding full-length, mature human BMP-2 [Wozney et al, Science 242:1528-1534 (1988)].

In this case the thioredoxin-BMP-2 fusion protein appeared in the insoluble cellular fraction when strain GI724 containing the expression vector was grown in medium containing tryptophan at 37°C. However, when the temperature of the growth medium was lowered to 20°C the fusion protein was found in the soluble cellular fraction.

**EXAMPLE 5 - THIOREDOXIN-SMALL PEPTIDE FUSION MOLECULES**

Native E. coli thioredoxin was expressed at high levels in E. coli using strain GI724 containing the same plasmid expression vector described in Example 3 deleted for nucleotides 2569-3129, and employing the growth and induction protocol outlined in Example 1. Under these conditions thioredoxin accumulated to approximately 10% of the total protein, all of it in the soluble cellular fraction.

Fig. 4 illustrates insertion of 13 amino acid residues encoding an enterokinase cleavage site into the active site loop of thioredoxin, between residues G<sub>34</sub> and P<sub>35</sub> of the thioredoxin protein sequence. The fusion protein containing this internal enterokinase site was expressed at levels equivalent to native thioredoxin, and was cleaved with an enterokinase treatment as outlined in Example 1 above. The fusion protein was found to be as stable as native thioredoxin to heat treatments, being resistant to a 10 minute incubation at 80°C as described in Example 4.

Below are listed twelve additional peptide insertions which



were also made into the active site loop of thioredoxin between G<sub>34</sub> and P<sub>35</sub>. The sequences are each 14 amino acid residues in length and are random in composition. Each of the thioredoxin fusion proteins containing these random insertions were made at levels comparable to native thioredoxin. All of them were found in the soluble cellular fraction. These peptides include the following sequences:

Pro-Leu-Gln-Arg-Ile-Pro-Pro-Gln-Ala-Leu-Arg-Val-Glu-Gly,  
 Pro-Arg-Asp-Cys-Val-Gln-Arg-Gly-Lys-Ser-Leu-Ser-Leu-Gly,  
 10 Pro-Met-Arg-His-Asp-Val-Arg-Cys-Val-Leu-His-Gly-Thr-Gly,  
 Pro-Gly-Val-Arg-Leu-Pro-Ile-Cys-Tyr-Asp-Asp-Ile-Arg-Gly,  
 Pro-Lys-Phe-Ser-Asp-Gly-Ala-Gln-Gly-Leu-Gly-Ala-Val-Gly,  
 Pro-Pro-Ser-Leu-Val-Gln-Asp-Asp-Ser-Phe-Glu-Asp-Arg-Gly,  
 Pro-Trp-Ile-Asn-Gly-Ala-Thr-Pro-Val-Lys-Ser-Ser-Ser-Gly,  
 15 Pro-Ala-His-Arg-Phe-Arg-Gly-Gly-Ser-Pro-Ala-Ile-Phe-Gly,  
 Pro-Ile-Met-Gly-Ala-Ser-His-Gly-Glu-Arg-Gly-Pro-Glu-Gly,  
 Pro-Asp-Ser-Leu-Arg-Arg-Arg-Glu-Gly-Phe-Gly-Leu-Leu-Gly,  
 Pro-Ser-Glu-Tyr-Pro-Gly-Leu-Ala-Thr-Gly-His-His-Val-Gly,  
 and Pro-Leu-Gly-Val-Leu-Gly-Ser-Ile-Trp-Leu-Glu-Arg-Gln-Gly.

20 The inserted sequences contained examples that were both hydrophobic and hydrophilic, and examples that contained cysteine residues. It appears that the active-site loop of thioredoxin can tolerate a wide variety of peptide insertions resulting in soluble fusion proteins. Standard procedures can be used to  
 25 purify these loop "inserts".

#### EXAMPLE 6 - HUMAN INTERLEUKIN-6

Human interleukin-6 (IL-6) was expressed at high levels in E. coli as a thioredoxin fusion protein using an expression  
 30 vector similar to modified pALtrxA/EK/IL11Δ Pro-581 described in Example 3 above. To express a thioredoxin-IL-6 fusion the DNA sequence in modified pALtrxA/EK/IL11Δ Pro-581 encoding human IL-11 (nucleotides 2599-3132) is replaced by the 561 nucleotide DNA sequence shown in Figur 6 encoding full-length, mature human IL-  
 35 6 [Hiran et al, Nature 324:73-76 (1986)]. The host strain and

expression protocol used for the production of thioredoxin-IL-6 fusion protein are as described in Example 1.

When the fusion protein was synthesized at 37°C, approximately 50% of it was found in the "inclusion body" or insoluble fraction. However all of the thioredoxin-IL-6 fusion protein, representing up to 10% of the total cellular protein, was found in the soluble fraction when the temperature of synthesis was lowered to 25°C.

10 EXAMPLE 7 - HUMAN MACROPHAGE COLONY STIMULATING FACTOR

Human Macrophage Colony Stimulating Factor (M-CSF) was expressed at high levels in E. coli as a thioredoxin fusion protein using the modified expression vector similar to pALtrxA/EK/IL11A Pro-581 described in Example 3 above.

15 The DNA sequence encoding human IL-11 in modified pALtrxA/EK/IL11A Pro-581 (nucleotides 2599-3135) is replaced by the 669 nucleotide DNA sequence shown in Fig. 7 encoding the first 223 amino acids of mature human M-CSF [G. G. Wong et al, Science 235:1504-1508 (1987)]. The host strain and expression  
20 protocol used for the production of thioredoxin-M-CSF fusion protein was as described in Example 2 above.

As was seen with the thioredoxin-IL-11 fusion protein, all of the thioredoxin-M-CSF fusion protein was found in the soluble cellular fraction, representing up to 10% of the total protein.

25

EXAMPLE 8 - RELEASE OF FUSION PROTEIN VIA OSMOTIC SHOCK OR FREEZE/THAW

To determine whether the fusions of heterologous proteins to thioredoxin according to this invention enable targeting to the host cell's adhesion sites and permit the release of the fusion  
30 proteins from the cell, the cells were exposed to simple osmotic shock and freeze/thaw procedures.

Cells overproducing wild-type E. coli thioredoxin, human thi redoxin, th E. coli thior doxin-MIP<sub>1α</sub> fusion or the E. coli  
35 thi r doxin-IL-11 fusion were us d in the following procedur s.

For an osmotic shock treatment, cells were resuspended at 2 A<sub>550</sub>/ml in 20 mM Tris-Cl pH 8.0/2.5 mM EDTA/20% w/v sucrose and kept cold on ice for 10 minutes. The cells were then pelleted by centrifugation (12,000 xg, 30 seconds) and gently resuspended in the same buffer as above but with sucrose omitted. After an additional 10 minute period on ice, to allow for the osmotic release of proteins, cells were re-pelleted by centrifugation (12,000 xg, 2 minutes) and the supernatant ("shockate") examined for its protein content. Wild-type E. coli thioredoxin and human thioredoxin were quantitatively released, giving "shockate" preparations which were >80% pure thioredoxin. More significantly >80% of the thioredoxin-MIP<sub>1</sub> and >50% of the thioredoxin-IL-11 fusion proteins were released by this osmotic treatment.

A simple freeze/thaw procedure produced similar results, releasing thioredoxin fusion proteins selectively, while leaving most of the other cellular proteins inside the cell. A typical freeze/thaw procedure entails resuspending cells at 2 A<sub>550</sub>/ml in 20 mM Tris-Cl pH 8.0/2.5 mM EDTA and quickly freezing the suspension in dry ice or liquid nitrogen. The frozen suspension is then allowed to slowly thaw before spinning out the cells (12,000 xg, 2 minutes) and examining the supernatant for protein.

Although the resultant "shockate" may require additional purification, the initial "shockate" is characterized by the absence of nucleic acid contaminants. Compared to an initial lysate, the purity of the "shockate" is significantly better, and does not require the difficult removal of DNA from bacterial lysates.

Thus, this release step can be substituted for the lysis step of Example 2. The supernatant obtained after centrifugation is then further purified in the manner disclosed in that Example.

Numerous modifications and variations of the present invention are included in the above-identified specification and are expected to be obvious to one of skill in the art. Such modifications and alterations to the compositions and processes

34

of the present invention are believed to be encompassed in the scope of the claims appended hereto.

## WHAT IS CLAIMED IS:

1. A DNA sequence encoding a fusion protein, said sequence comprising DNA encoding a thioredoxin-like protein fused to a DNA sequence encoding a selected heterologous protein.
2. A DNA sequence of claim 1 wherein said DNA encoding said thioredoxin-like protein comprises the amino terminus of said fusion protein.
3. A DNA sequence of claim 1 wherein said DNA encoding said thioredoxin-like protein comprises the carboxyl terminus of said fusion protein.
4. A DNA sequence of claim 1, 2 or 3 wherein said DNA encoding said thioredoxin-like protein is selected from the group consisting of E. coli thioredoxin and human thioredoxin.
5. A DNA sequence of claim 1, 2 or 3 wherein said DNA encoding said selected protein is selected from the group consisting of IL-11, IL-6, Macrophage Inhibitory Protein 1 $\alpha$  and Bone Morphogenic Protein 2.
6. A DNA sequence of claim 1, 2 or 3 additionally comprising a linker DNA sequence fused between said DNA encoding said thioredoxin-like protein and said DNA encoding said selected heterologous protein.
7. A plasmid DNA molecule comprising a DNA sequence of claims 1-6, said sequence being under the control of a suitable expression control sequence capable of directing the expression of a fusion protein in a selected host cell.
8. An E. coli host cell transformed with, or having integrated into the genome thereof, a plasmid of claim 7.

9. A method of making a selected heterologous protein comprising
- (a) culturing in a culture medium under suitable conditions a host cell of claim 8;
  - (b) recovering the fusion protein produced thereby from said culture medium;
  - (c) cleaving said selected heterologous protein from said fusion protein and
  - (d) isolating said selected heterologous protein.
10. IL-11 protein produced by the method of claim 9.
11. Use of thioredoxin in the method of claim 9.

1/12

## FIGURE 1A

pALtrxA/EK/IL11A Pro-581

GACGAAAGGG CCTCGTGATA CGCCTATTTT TATAGGTTAA	40
TGTCATGATA ATAATGGTTT CTAGACGTC AGGTGGCACT	80
TTTCGGGGAA ATGTGCGCGG AACCCTATT TGTTTATTTT	120
TCTAAATACA TTCAAATATG TATCCGCTCA TGAGACAATA	160
ACCCTGATAA ATGCTTCAAT AATATTGAAA AAGGAAGAGT	200
ATGAGTATTC AACATTTCCG TGTCGCCCTT ATTCCCTTTT	240
TTGCGGCATT TTGCCTTCCT GTTTTGTCTC ACCCAGAAAC	280
GCTGGTGAAA GTAAAAGATG CTGAAGATCA GTTGGGTGCA	320
CGAGTGGGTT ACATCGAACT GGATCTCAAC AGCGGTAAGA	360
TCCTTGAGAG TTTTCGCCCC GAAGAACGTT TTCCAATGAT	400
GAGCACTTTT AAAGTTCTGC TATGTGGCGC GGTATTATCC	440
CGTATTGACG CCGGGCAAGA GCAACTCGGT CGCCGCATAC	480
ACTATTCTCA GAATGACTTG GTTGAGTACT CACCAGTCAC	520
AGAAAAGCAT CTTACGGATG GCATGACAGT AAGAGAATTA	560
TGCAGTGCTG CCATAACCAT GAGTGATAAC ACTGCGGCCA	600
ACTTACTTCT GACAACGATC GGAGGACCGA AGGAGCTAAC	640
CGCTTTTTTG CACAACATGG GGGATCATGT AACTCGCCTT	680
GATCGTTGGG AACCGGAGCT GAATGAAGCC ATACCAAACG	720
ACGAGCGTGA CACCACGATG CCTGTAGCAA TGGCAACAAC	760
GTTGCGCAAA CTATTAACTG GCGAACTACT TACTCTAGCT	800
TCCCGGCAAC AATTAATAGA CTGGATGGAG GCGGATAAAG	840
TTGCAGGACC ACTTCTGCGC TCGGCCCTTC CGGCTGGCTG	880
GTTTATTGCT GATAAATCTG GAGCCGGTGA GCGTGGGTCT	920
CGCGGTATCA TTGCAGCACT GGGGCCAGAT GGTAAGCCCT	960
CCCGTATCGT AGTTATCTAC ACGACGGGGA GTCAGGCAAC	1000
TATGGATGAA CGAAATAGAC AGATCGCTGA GATAGGTGCC	1040

SUBSTITUTE SHEET

2/12

## FIGURE 1B

TCACTGATTA AGCATTGGTA ACTGTCAGAC CAAGTTTACT	1080
CATATATACT TTAGATTGAT TTAAAACTTC ATTTTAAATT	1120
TAAAAGGATC TAGGTGAAGA TCCTTTTTGA TAATCTCATG	1160
ACCAAAATCC CTTAACGTGA GTTTTCGTTC CACTGAGCGT	1200
CAGACCCCGT AGAAAAGATC AAAGGATCTT CTTGAGATCC	1240
TTTTTTTCTG CGCGTAATCT GCTGCTTGCA AACAAAAAAA	1280
CCACCGCTAC CAGCGGTGGT TTGTTTGCCG GATCAAGAGC	1320
TACCAACTCT TTTTCCGAAG GTAAC TGGCT TCAGCAGAGC	1360
GCAGATACCA AATACTGTCC TTCTAGTGTA GCCGTAGTTA	1400
GGCCACCACT TCAAGAACTC TGTAGCACCG CCTACATACC	1440
TCGCTCTGCT AATCCTGTTA CCAGTGGCTG CTGCCAGTGG	1480
CGATAAGTCG TGTCTTACCG GGT TGGACTC AAGACGATAG	1520
TTACCGGATA AGGCGCAGCG GTCGGGCTGA ACGGGGGGTT	1560
CGTGCACACA GCCCAGCTTG GAGCGAACGA CCTACACCGA	1600
ACTGAGATAC CTACAGCGTG AGCATTGAGA AAGCGCCACG	1640
CTTCCCGAAG GGAGAAAGGC GGACAGGTAT CCGGTAAGCG	1680
GCAGGGTCGG AACAGGAGAG CGCAGCAGGG AGCTTCCAGG	1720
GGGAAACGCC TGGTATCTTT ATAGTCCTGT CGGGTTTCGC	1760
CACCTCTGAC TTGAGCGTCG ATTTTTGTGA TGCTCGTCAG	1800
GGGGGCGGAG CCTATGGAAA AACGCCAGCA ACGCGGCCTT	1840
TTTACGGTTC CTGGCCTTTT GCTGGCCTTT TGCTCACATG	1880
TTCTTTTCTG CGTTATCCCC TGATTCTGTG GATAACCGTA	1920
TTACCGCCTT TGAGTGAGCT GATACCGCTC GCCGCAGCCG	1960
AACGACCGAG CGCAGCGAGT CAGTGAGCGA GGAAGCGGAA	2000
GAGCGCCCAA TACGCAAACC GCCTCTCCCC GCGCGTTGGC	2040
CGATTCATTA ATGCAGAATT GATCTCTCAC CTACCAAACA	2080
ATGCCCCCCT GCAAAAAATA AATTCATATA AAAACATAC	2120

SUBSTITUTE SHEET



3/12

## FIGURE 1C

AGATAACCAT CTGCGGTGAT AAATTATCTC TGGCGGTGTT	2160
GACATAAATA CCACTGGCGG TGATACTGAG CACATCAGCA	2200
GGACGCACTG ACCACCATGA ATTCAAGAAG GAGATATACA	2240
T ATG AGC GAT AAA ATT ATT CAC CTG ACT GAC GAC Met Ser Asp Lys Ile Ile His Leu Thr Asp Asp 1 5 10	2274
AGT TTT GAC ACG GAT GTA CTC AAA GCG GAC GGG Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly 15 20	2307
GCG ATC CTC GTC GAT TTC TGG GCA GAG TGG TGC Ala Ile Leu Val Asp Phe Trp Ala Glu Trp Cys 25 30	2340
GGT CCG TGC AAA ATG ATC GCC CCG ATT CTG GAT Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp 35 40	2373
GAA ATC GCT GAC GAA TAT CAG GGC AAA CTG ACC Glu Ile Ala Asp Glu Tyr Gln Gly Lys Leu Thr 45 50 55	2406
GTT GCA AAA CTG AAC ATC GAT CAA AAC CCT GGC Val Ala Lys Leu Asn Ile Asp Gln Asn Pro Gly 60 65	2439
ACT GCG CCG AAA TAT GGC ATC CGT GGT ATC CCG Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile Pro 70 75	2472
ACT CTG CTG CTG TTC AAA AAC GGT GAA GTG GCG Thr Leu Leu Leu Phe Lys Asn Gly Glu Val Ala 80 85	2505
GCA ACC AAA GTG GGT GCA CTG TCT AAA GGT CAG Ala Thr Lys Val Gly Ala Leu Ser Lys Gly Gln 90 95	2538
TTG AAA GAG TTC CTC GAC GCT AAC CTG GCC GGT Leu Lys Glu Phe Leu Asp Ala Asn Leu Ala Gly 100 105 110	2571
TCT GGT TCT GGT GAT GAC GAT GAC AAA GGT CCA Ser Gly Ser Gly Asp Asp Asp Lys Gly Pro 115 120	2604

SUBSTITUTE SHEET

4/12

## FIGURE 1D

CCA CCA GGT CCA CCT CGA GTT TCC CCA GAC CCT Pro Pro Gly Pro Pro Arg Val Ser Pro Asp Pro 125 130	2637
CGG GCC GAG CTG GAC AGC ACC GTG CTC CTG ACC Arg Ala Glu Leu Asp Ser Thr Val Leu Leu Thr 135 140	2670
CGC TCT CTC CTG GCG GAC ACG CGG CAG CTG GCT Arg Ser Leu Leu Ala Asp Thr Arg Gln Leu Ala 145 150	2703
GCA CAG CTG AGG GAC AAA TTC CCA GCT GAC GGG Ala Gln Leu Arg Asp Lys Phe Pro Ala Asp Gly 155 160 165	2736
GAC CAC AAC CTG GAT TCC CTG CCC ACC CTG GCC Asp His Asn Leu Asp Ser Leu Pro Thr Leu Ala 170 175	2769
ATG AGT GCG GGG GCA CTG GGA GCT CTA CAG CTC Met Ser Ala Gly Ala Leu Gly Ala Leu Gln Leu 180 185	2802
CCA GGT GTG CTG ACA AGG CTG CGA GCG GAC CTA Pro Gly Val Leu Thr Arg Leu Arg Ala Asp Leu 190 195	2835
CTG TCC TAC CTG CGG CAC GTG CAG TGG CTG CGC Leu Ser Tyr Leu Arg His Val Gln Trp Leu Arg 200 205	2868
CGG GCA GGT GGC TCT TCC CTG AAG ACC CTG GAG Arg Ala Gly Gly Ser Ser Leu Lys Thr Leu Glu 210 215 220	2901
CCC GAG CTG GGC ACC CTG CAG GCC CGA CTG GAC Pro Glu Leu Gly Thr Leu Gln Ala Arg Leu Asp 225 230	2934
CGG CTG CTG CGC CGG CTG CAG CTC CTG ATG TCC Arg Leu Leu Arg Arg Leu Gln Leu Leu Met Ser 235 240	2967
CGC CTG GCC CTG CCC CAG CCA CCC CCG GAC CCG Arg Leu Ala Leu Pro Gln Pro Pro Pro Asp Pro 245 250	3000
CCG GCG CCC CCG CTG GCG CCC CCC TCC TCA GCC Pro Ala Pro Pro Leu Ala Pro Pro Ser Ser Ala 255 260	3033

SUBSTITUTE SHEET

5/12

## FIGURE 1E

TGG GGG GGC ATC AGG GCC GCC CAC GCC ATC CTG Trp Gly Gly Ile Arg Ala Ala His Ala Ile Leu 265 270 275	3066
GGG GGG CTG CAC CTG ACA CTT GAC TGG GCC GTG Gly Gly Leu His Leu Thr Leu Asp Trp Ala Val 280 285	3099
AGG GGA CTG CTG CTG CTG AAG ACT CGG CTG TGA Arg Gly Leu Leu Leu Leu Lys Thr Arg Leu 290 295	3132
AAGCTTATCG ATACCGTCGA CCTGCAGTAA TCGTACAGGG	3172
TAGTACAAAT AAAAAAGGCA CGTCAGATGA CGTGCCTTTT	3212
TTCTTGTGAG CAGTAAGCTT GGCACITGGCC GTCGTTTTAC	3252
AACGTCGTGA CTGGGAAAAC CCTGGCGTTA CCCAACTTAA	3292
TCGCCTTGCA GCACATCCCC CTTTCGCCAG CTGGCGTAAT	3332
AGCGAAGAGG CCCGCACCGA TCGCCCTTCC CAACAGTTGC	3372
GCAGCCTGAA TGGCGAATGG CGCCTGATGC GGTATTTTCT	3412
CCTTACGCAT CTGTGCGGTA TTTCACACCG CATATATGGT	3452
GCACTCTCAG TACAATCTGC TCTGATGCCG CATAGTTAAG	3492
CCAGCCCCGA CACCCGCCAA CACCCGCTGA CGCGCCCTGA	3532
CGGGCTTGTC TGCTCCCGGC ATCCGCTTAC AGACAAGCTG	3572
TGACCGTCTC CGGGAGCTGC ATGTGTCAGA GGTTTTCACC	3612
GTCATCACCG AAACGCGCGA	3632

6/12

FIGURE 2

## MIP-1a

GCA	CCA	CTT	GCT	GCT	GAC	ACG	CCG	ACC	GCC	TGC	TGC	36
Ala	Pro	Leu	Ala	Ala	Asp	Thr	Pro	Thr	Ala	Cys	Cys	
1				5					10			
TTC	AGC	TAC	ACC	TCC	CGA	CAG	ATT	CCA	CAG	AAT	TTC	72
Phe	Ser	Tyr	Thr	Ser	Arg	Gln	Ile	Pro	Gln	Asn	Phe	
		15					20					
ATA	GCT	GAC	TAC	TTT	GAG	ACG	AGC	AGC	CAG	TGC	TCC	109
Ile	Ala	Asp	Tyr	Phe	Glu	Thr	Ser	Ser	Gln	Cys	Ser	
25					30					35		
AAG	CCC	AGT	GTC	ATC	TTC	CTA	ACC	AAG	AGA	GGC	CGG	145
Lys	Pro	Ser	Val	Ile	Phe	Leu	Thr	Lys	Arg	Gly	Arg	
			40					45				
CAG	GTC	TGT	GCT	GAC	CCC	AGT	GAG	GAG	TGG	GTC	CAG	181
Gln	Val	Cys	Ala	Asp	Pro	Ser	Glu	Glu	Trp	Val	Gln	
	50					55					60	
AAA	TAC	GTC	AGT	GAC	CTG	GAG	CTG	AGT	GCC	TAA		214
Lys	Thr	Val	Ser	Asp	Leu	Glu	Leu	Ser	Ala			
				65					70			

7/12

**FIGURE 3**

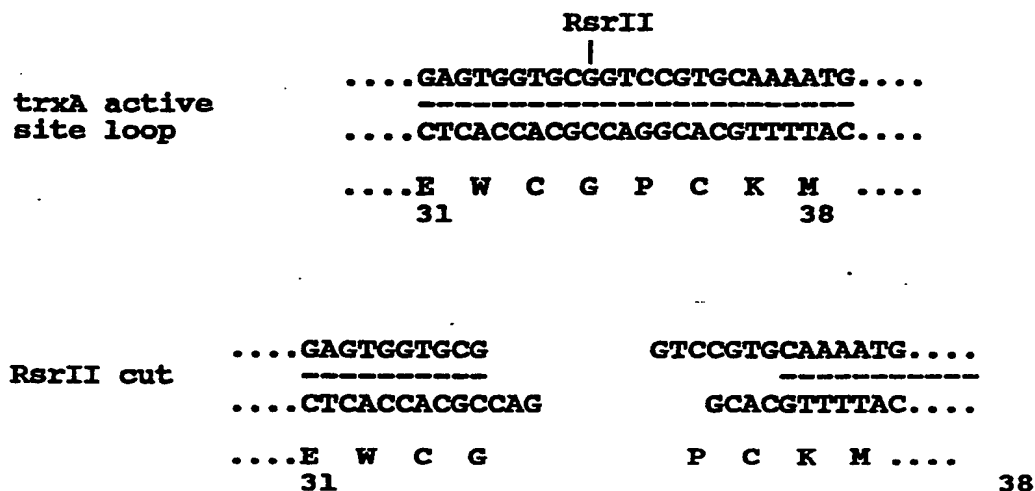
**BMP-2**

CAA Gln 1	GCT Ala	AAA Lys	CAT His	AAA Lys 5	CAA Gln	CGT Arg	AAA Lys	CGT Arg	CTG Leu 10	AAA Lys	TCT Ser	36
AGC Ser	TGT Cys	AAG Lys 15	AGA Arg	CAC His	CCT Pro	TTG Leu	TAC Tyr 20	GTG Val	GAC Asp	TTC Phe	AGT Ser	72
GAC Asp 25	GTG Val	GGG Gly	TGG Trp	AAT Asn 30	GAC Asp 30	TGG Trp	ATT Ile	GTG Val	GCT Ala	CCC Pro 35	CCG Pro	109
GGG Gly	TAT Tyr	CAC His 40	GCC Ala	TTT Phe 40	TAC Tyr	TGC Cys	CAC His 45	GGA Gly 45	GAA Glu	TGC Cys	CCT Pro	145
TTT Phe 50	CCT Pro 50	CTG Leu	GCT Ala	GAT Asp	CAT His 55	CTG Leu 55	AAC Asn	TCC Ser	ACT Thr	AAT Asn 60	CAT His 60	181
GCC Ala	ATT Ile	GTT Val	CAG Gln 65	ACG Thr 65	TTG Leu	GTC Val	AAC Asn	TCT Ser	GTT Val 70	AAC Asn	TCT Ser	217
AAG Lys	ATT Ile 75	CCT Pro 75	AAG Lys	GCA Ala	TGC Cys	TGT Cys 80	GTC Val 80	CCG Pro	ACA Thr	GAA Glu	CTC Leu	253
AGT Ser 85	GCT Ala	ATC Ile	TCG Ser	ATG Met 90	CTG Leu 90	TAC Tyr	CTT Leu	GAC Asp	GAG Glu	AAT Asn 95	GAA Glu	289
AAG Lys	GTT Val	GTA Val 100	TTA Leu 100	AAG Lys	AAC Asn	TAT Tyr	CAG Gln 105	GAC Asp 105	ATG Met	GTT Val	GTG Val	325
GAG Glu 110	GGT Gly 110	TGT Cys	GGG Gly	TGT Cys	CGC Arg	TAG						346

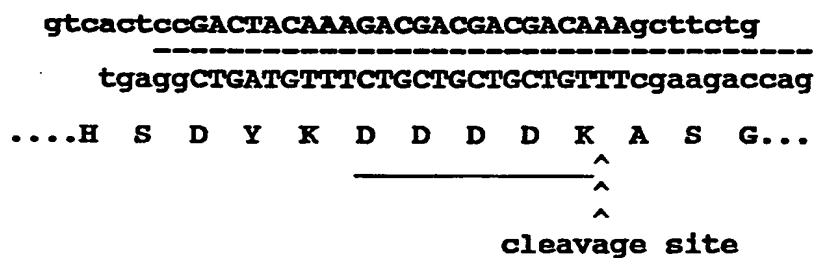
8/12

FIGURE 4

INSERTION OF AN ENTEROKINASE SITE INTO  
THE ACTIVE-SITE LOOP OF E.COLI THIOREDOXIN (trxA)



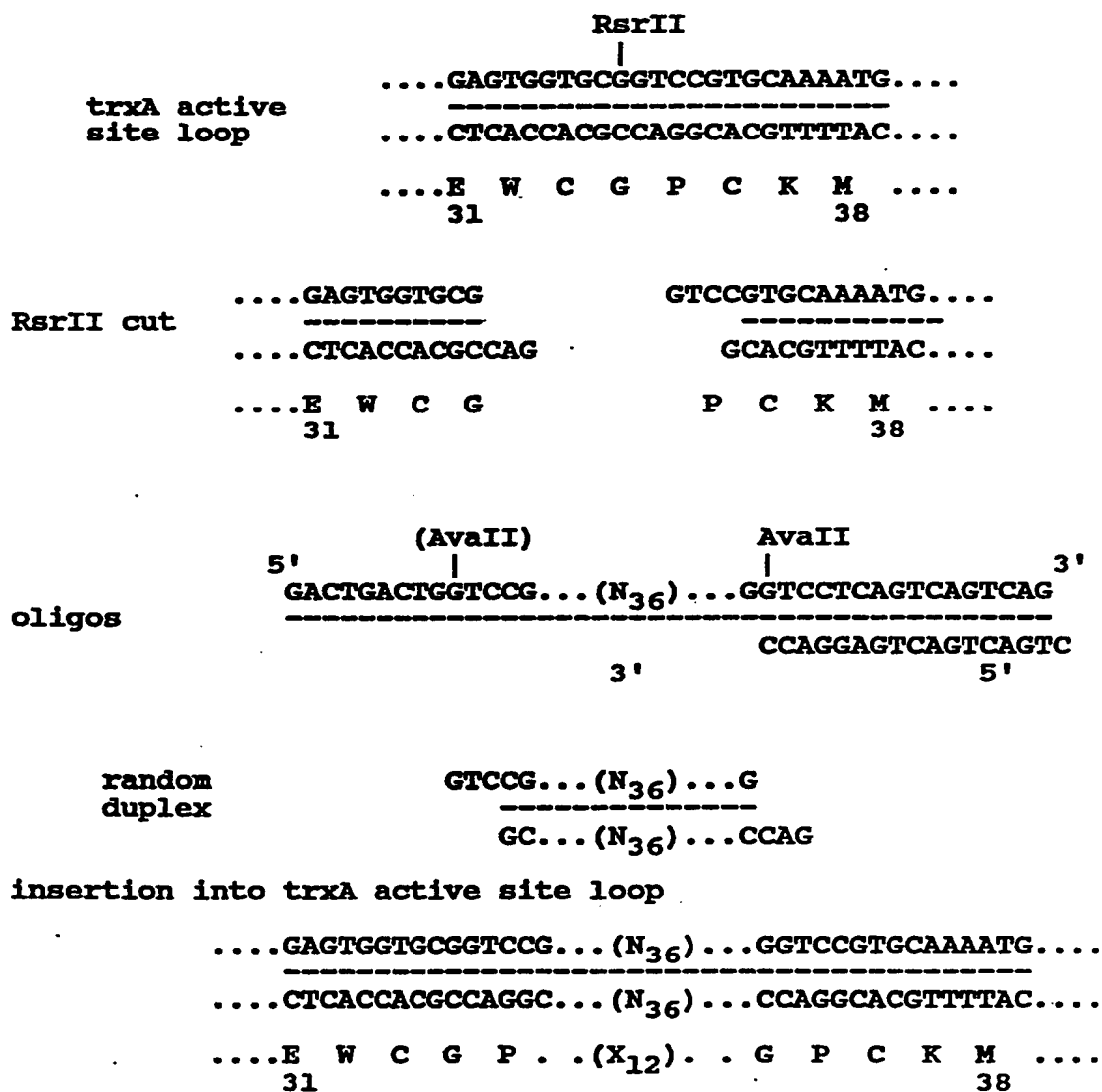
Enterokinase site  
(13 residues)



9/12

**FIGURE 5**

# RANDOM PEPTIDE INSERTIONS INTO THE ACTIVE-SITE LOOP OF E. COLI THIOREDOXIN (trxA)



10/12

**FIGURE 6**

IL-6

<div style="display: flex; justify-content: space-between; margin: 0;"> <span>5</span> <span>10</span> </div>													39
ATG	GCT	CCA	GTA	CCT	CCA	GGT	GAA	GAT	TCT	AAA	GAT	GTA	
Met	Ala	Pro	Val	Pro	Pro	Gly	Glu	Asp	Ser	Lys	Asp	Val	
<div style="display: flex; justify-content: space-between; margin: 0;"> <span>15</span> <span>20</span> <span>25</span> </div>													78
GCC	GCC	CCA	CAC	AGA	CAG	CCA	CTC	ACC	TCT	TCA	GAA	CGA	
Ala	Ala	Pro	His	Arg	Gln	Pro	Leu	Thr	Ser	Ser	Glu	Arg	
<div style="display: flex; justify-content: space-between; margin: 0;"> <span>30</span> <span>35</span> </div>													117
ATT	GAC	AAA	CAA	ATT	CGG	TAC	ATC	CTC	GAC	GGC	ATC	TCA	
Ile	Asp	Lys	Gln	Ile	Arg	Tyr	Ile	Leu	Asp	Gly	Ile	Ser	
<div style="display: flex; justify-content: space-between; margin: 0;"> <span>40</span> <span>45</span> <span>50</span> </div>													156
GCC	CTG	AGA	AAG	GAG	ACA	TGT	AAC	AAG	AGT	AAC	ATG	TGT	
Ala	Leu	Arg	Lys	Glu	Thr	Cys	Asn	Lys	Ser	Asn	Met	Cys	
<div style="display: flex; justify-content: space-between; margin: 0;"> <span>55</span> <span>60</span> </div>													195
GAA	AGC	AGC	AAA	GAG	GCA	CTG	GCA	GAA	AAC	AAC	CTG	AAC	
Glu	Ser	Ser	Lys	Glu	Ala	Leu	Ala	Glu	Asn	Asn	Leu	Asn	
<div style="display: flex; justify-content: space-between; margin: 0;"> <span>65</span> <span>70</span> <span>75</span> </div>													234
CTT	CCA	AAG	ATG	GCT	GAA	AAA	GAT	GGA	TGC	TTC	CAA	TCT	
Leu	Pro	Lys	Met	Ala	Glu	Lys	Asp	Gly	Cys	Phe	Gln	Ser	
<div style="display: flex; justify-content: space-between; margin: 0;"> <span>80</span> <span>85</span> <span>90</span> </div>													273
GGA	TTC	AAT	GAG	GAG	ACT	TGC	CTG	GTG	AAA	ATC	ATC	ACT	
Gly	Phe	Asn	Glu	Glu	Thr	Cys	Leu	Val	Lys	Ile	Ile	Thr	
<div style="display: flex; justify-content: space-between; margin: 0;"> <span>95</span> <span>100</span> </div>													312
GGT	CTT	TTG	GAG	TTT	GAG	GTA	TAC	CTA	GAG	TAC	CTC	CAG	
Gly	Leu	Leu	Glu	Phe	Glu	Val	Tyr	Leu	Glu	Thr	Leu	Gln	
<div style="display: flex; justify-content: space-between; margin: 0;"> <span>105</span> <span>110</span> <span>115</span> </div>													351
AAC	AGA	TTT	GAG	AGT	AGT	GAG	GAA	CAA	GCC	AGA	GCT	GTG	
Asn	Arg	Phe	Glu	Ser	Ser	Glu	Glu	Gln	Ala	Arg	Ala	Val	
<div style="display: flex; justify-content: space-between; margin: 0;"> <span>120</span> <span>125</span> </div>													390
CAG	ATG	AGT	ACA	AAA	GTC	CTG	ATC	CAG	TTC	CTG	CAG	AAA	
Gln	Met	Ser	Thr	Lys	Val	Leu	Ile	Gln	Phe	Leu	Gln	Lys	



11/12

FIGURE 6 (c ntinued)

130		140		150		
AAG GCA AAG AAT CTA GAT GCA ATA ACC ACC CCT GAC CCA	429					
Lys Ala Lys Asn Leu Asp Ala Ile Thr Thr Pro Asp Pro						
	155		160			
ACC ACA AAT GCC AGC CTG CTG ACG AAG CTG CAG GCA CAG	468					
Thr Thr Asn Ala Ser Leu Leu Thr Lys Leu Gln Ala Gln						
	170		175		180	
AAC CAG TGG CTG CAG GAC ATG ACA ACT CAT CTC ATT CTG	507					
Asn Gln Trp Leu Gln Asp Met Thr Thr His Leu Ile Leu						
	185		190			
CGC AGC TTT AAG GAG TTC CTG CAG TCC AGC CTG AGG GCT	546					
Arg Ser Phe Lys Glu Phe Leu Gln Ser Ser Leu Arg Ala						
	195					
CTT CGG CAA ATG TAG	561					
Leu Arg Gln Met *						

12/12

## FIGURE 7

1 GAAGAAGTTT CTGAATATTG TAGCCACATG ATTGGGAGTG GACACCTGCA  
51 GTCTCTGCAG CGGCTGATTG ACAGTCAGAT GGAGACCTCG TGCCAAATTA  
101 CATTTGAGTT TGTAGACCAG GAACAGTTGA AAGATCCAGT GTGCTACCTT  
151 AAGAAGGCAT TTCTCCTGGT ACAAGACATA ATGGAGGACA CCATGCGCTT  
201 CAGAGATAAC ACCCCCAATG CCATCGCCAT TGTGCAGCTG CAGGAACTCT  
251 CTTTGAGGCT GAAGAGCTGC TTCACCAAGG ATTATGAAGA GCATGACAAG  
301 GCCTGCGTCC GAACTTTCTA TGAGACACCT CTCCAGTTGC TGGAGAAGGT  
351 CAAGAATGTC TTFAATGAAA CAAAGAATCT CCTTGACAAG GACTGGAATA  
401 TTTTCAGCAA GAACTGCAAC AACAGCTTTG CTGAATGCTC CAGCCAAGAT  
451 GTGGTGACCA AGCCTGATTG CAACTGCCTG TACCCCAAAG CCATCCCTAG  
501 CAGTGACCCG GCCTCTGTCT CCCCTCATCA GCCCCTCGCC CCCTCCATGG  
551 CCCCTGTGGC TGGCTTGACC TGGGAGGACT CTGAGGGAAC TGAGGGCAGC  
601 TCCCTCTTGC CTGGTGAGCA GCCCCTGCAC ACAGTGGATC CAGGCAGTGC  
651 CAAGCAGCGG CCACCCAGG

# INTERNATIONAL SEARCH REPORT

International Application No. PCT/US92/00944

<b>I. CLASSIFICATION F SUBJECT MATTER</b> (if several classification symbols apply, indicate all) <sup>3</sup>		
According to International Patent Classification (IPC) or to both National Classification and IPC		
IPC (5): C 12N 15/70, 1/21, C07H 21/04; C07K 13/00; C12P 21/02		
US CL : 435/69.1, 252.33, 320.1; 530/351; 536/27		
<b>II. FIELDS SEARCHED</b>		
Minimum Documentation Searched <sup>4</sup>		
Classification System	Classification Symbols	
U.S.	435/69.1, 69.7, 240.1, 243, 252.33, 320.1; 530/351; 536/27; 935/10, 27, 72	
Documentation Searched other than Minimum Documentation to the extent that such Documents are included in the Fields Searched <sup>5</sup>		
Please See Attached Sheet.		
<b>III. DOCUMENTS CONSIDERED TO BE RELEVANT<sup>14</sup></b>		
Category <sup>*</sup>	Citation of Document <sup>18</sup> with indication, where appropriate, of the relevant passages <sup>17</sup>	Relevant to Claim No. <sup>18</sup>
X/Y	Proceedings of the National Academy of Sciences, U.S.A. Volume 87, issued November 1990, N. Wakasugi et al. "Adult T-cell leukemia-derived factor/thioredoxin, produced by both human T-lymphotrophic virus tyoe I- and Epstein-Barr virus-transformed lymphocytes, acts as an autocrine growth factor and synergizes with interleukin 1 and interleukin 2", pages 8282-8286, see introduction.	1-4, 6-9 / 5, 11
X/Y,P	US,A, 5,011,772 (Recsei) 30 April 1991, see Example 4.	1-4, 6-9/ 5,11
X/Y	Proceedings of the National Academy of Sciences, U.S.A., Volume 87, issued October 1990, S.R. Paul et al, "Molecular cloning of a cDNA encoding interleukin 11, a stromal cell-derived lymphopoietic and hematopoietic cytokine", pages 7512-7516, see entire document.	9/5
<p><sup>*</sup> Special categories of cited documents:<sup>16</sup></p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>"&amp;" document member of the same patent family</p>		
<b>IV. CERTIFICATION</b>		
Date of the Actual Completion of the International Search <sup>2</sup>		Date of Mailing of this International Search Report <sup>2</sup>
22 APRIL 1992		13 MAY 1992
International Searching Authority <sup>1</sup>		Signature of Authorized Office <sup>20</sup>
ISA/US		<i>Andree Reiman</i> Gabriele E. Bugaisky

## FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

V. ☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE<sup>1</sup>

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. ☐ Claim numbers , because they relate to subject matter (1) not required to be searched by this Authority, namely:

2. ☐ Claim numbers , because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out (1), specifically:

3. ☐ Claim numbers , because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. ☒ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING<sup>2</sup>

This International Searching Authority found multiple inventions in this international application as follows:

I. Claims 1-9 and 11, drawn to a first composition and a first method of use, classified in Class 435/69.1.

II. Claim 10, drawn to a second composition, classified in Class 530/351.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application. (Telephone Practice)

2. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Search Authority did not invite payment of any additional fee.

Remark on protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PREVIOUS SHEETS

II. FIELDS SEARCHED

Other Documents Searched:

APS; DIALOG; MEDLINE, BIOSIS, BIOTECH ABS; GENBANK, PIR. -search terms: interleukin 6, interleukin 11, IL-6, IL-11, fusion protein, fusion peptide, thioredoxin, chimeric protein, prrotein, chimera, macrophage inhibitory protein 1 $\alpha$ , MIP-1 $\alpha$ , bone morphogenic protein 2, BMP 2, interleukin 1, IL-1, glutaredoxin, protein disulfide isomerase, stabilization, stabilize, stability, accumulation, accumulate

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☒ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**